Influence of the event magnitude on the predictability of an extreme event

S. Hallerberg and H. Kantz

Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Strasse 38, D 01187 Dresden, Germany

We investigate the predictability of extreme events in time series. The focus of this work is to understand under which circumstances large events are better predictable than smaller events. Therefore we use a simple prediction algorithm based on precursory structures which are identified using the maximum likelihood principle. Using the receiver operator characteristic curve as a measure for the quality of predictions we find that the dependence on the event size is closely linked to the probability distribution function of the underlying stochastic process. We evaluate this dependence on the probability distribution function analytically and numerically. If we assume that the optimal precursory structures are used to make the predictions, we find that large increments are better predictable if the underlying stochastic process has a Gaussian probability distribution function, whereas larger increments are harder to predict if the underlying probability distribution function has a power-law tail. In the case of an exponential distribution function we find no significant dependence on the event size. Furthermore we compare these results with predictions of increments in correlated data, namely, velocity increments of a free jet flow. The velocity increments in the free jet flow are in dependence on the time scale either asymptotically Gaussian or asymptotically exponential distributed. The numerical results for predictions within free jet data are in good agreement with the previous analytical considerations for random numbers.

I. INTRODUCTION

Systems with a complex time evolution, which generate a great impact event from time to time, are ubiquitous. Examples include fluctuations of prices for financial assets in economy with rare market crashes, electrical activity of human brain with rare epileptic seizures, seismic activity of the earth with rare earthquakes, changing weather conditions with rare disastrous storms, and also fluctuations of online diagnostics of technical machinery and networks with rare breakdowns or blackouts. Due to the complexity of the systems mentioned, a complete modeling is usually impossible, either due to the huge number of degrees of freedom involved, or due to a lack of precise knowledge about the governing equations. This is why one applies the framework of prediction via precursory structures for such cases. The typical application for prediction with precursory structures is a prediction of an event which occurs in the very near future, i.e., on short time scales compared to the lifetime of the system. A classical example for the search for precursory structures is the prediction of earthquakes [1]. A more recently studied example is the short term prediction of strong turbulent wind gusts, which can destroy wind turbines [2,3].

In a previous work [4], we studied the quality of predictions analytically via precursory structures for increments in an AR(1) process and numerically in a long-range correlated ARMA process. The long-range correlations did not alter the general findings for Gaussian processes, namely, that larger events are better predictable.

Furthermore we found other works which report the same effect for earthquake prediction [5], prediction of avalances in models which display self organized criticality [6] and in multiagent games [7]. In this contribution, we investigate the influence of the probability distribution function (PDF) of the noise term in detail by using not only Gaussian, but also

exponential and power-law distributed noise. This approach is also motivated by the book of Egans [8] which explains that receiver operating characteristic (ROC) curves obtained in signal detection problems can be ordered families of functions in dependence on a parameter. We are now interested in learning how the behavior of these families of functions depends on the event size and the distribution of the stochastic process, if the ROC curve is used for evaluating the quality of predictions.

After defining the prediction scheme in Sec. II A and the method for measuring the quality of a prediction in Sec. II B, we explain in Sec. II C how to consider the influence on the event size. In Sec. II D we formulate a constraint, which has to be fulfilled in order to find a better predictability of larger (smaller) events. In the next section, we apply this constraint to compare the quality of predictions of large increments within Gaussian (Sec. III A), exponential distributed (Sec. III B), and power-law distributed independent identically distributed (i.i.d.) random numbers (Sec. III C). We study the prediction of increments in free jet data in Sec. IV. Conclusions appear in Sec. V.

II. DEFINITIONS AND SETUP

The considerations in this section are made for a time series [9,10], i.e., a set of measurements x_n at discrete times t_n, where $t_n = t_0 + n\Delta$ with a sampling interval Δ and $n \in \mathbb{N}$. The recording should contain sufficiently many extreme events so that we are able to extract statistical information about them. We also assume that the event of interest can be identified on the basis of the observations, e.g., by the value of the observation function exceeding some threshold, by a sudden increase, or by its variance exceeding some threshold. We express the presence (absence) of an event by using a binary variable Y_{n+1}.

$$Y_{n+1} = \begin{cases} 1 & \text{an event occurred at time } n+1 \\ 0 & \text{no event occurred at time } n+1. \end{cases} \quad (1)$$

A. Choice of the precursor

When we consider prediction via precursory structures (*precursors*, or *predictors*), we are typically in a situation where we assume that the dynamics of the system under study has both, a deterministic and a stochastic part. The deterministic part allows one to assume that there is a relation between the event and its precursory structure which we can use for predictive purposes. However, if the dynamic of the system was fully deterministic there would be no need to predict via precursory structures, but we could exploit our knowledge about the dynamical system as it is done, e.g., in weather forecasting.

In this contribution we focus on the influence of the stochastic part of the dynamics and assume therefore a very simple deterministic correlation between event and precursor. The presence of this stochastic part determines that we cannot expect the precursor to precede *every* individual event. That is why we define a precursor in this context as a data structure which is *typically* preceding an event, allowing deviations from the given structure, but also allowing events without preceding structure.

For reasons of simplicity the following considerations are made for precursors in real space, i.e., structures in the time series. However, there is no reason not to apply the same ideas for precursory structures, which live in phase space.

In order to predict an event Y_{n+1} occurring at the time $(n+1)$ we compare the last k observations, to which we will refer as the *precursory variable*

$$\mathbf{x}_{(n-k+1,n)} = (x_{n-k+1}, x_{n-k+2}, \ldots, x_{n-1}, x_n), \quad (2)$$

with a specific precursory structure

$$\mathbf{x}^{pre} = (x_{n-k+1}^{pre}, x_{n-k+2}^{pre}, \ldots, x_{n-1}^{pre}, x_n^{pre}). \quad (3)$$

Once the precursory structure $\mathbf{x}_{pre}$ is determined, we give an alarm for an event $Y_{n+1}=1$ when we find the precursory variable $\mathbf{x}_{(n-k+1,n)}$ inside the volume

$$V^{pre}(\delta, \mathbf{x}^{pre}) = \prod_{j=n-k+1}^{n} \left(x_j^{pre} - \frac{\delta}{2}, x_j^{pre} + \frac{\delta}{2} \right). \quad (4)$$

There are different strategies to identify suitable precursory structures. We choose the precursor via maximizing a conditional probability which we refer to as the *likelihood* [11,25]. The likelihood

$$L(Y_{n+1} = 1 | \mathbf{x}_{(n-k+1,n)}) = \frac{j((Y_{n+1} = 1), \mathbf{x}_{(n-k+1,n)})}{\rho(\mathbf{x}_{(n-k+1,n)})}. \quad (5)$$

provides the probability that an event $Y_{n+1}=1$ follows the precursor $\mathbf{x}_{(n-k+1,n)}$. It can be calculated numerically by using the joint PDF $j((Y_{n+1}=1), \mathbf{x}_{(n-k+1,n)})$. Our prediction strategy consists of determining those values of each component x_i of $\mathbf{x}_{(n-k+1,n)}$ for which the likelihood is maximal.

This strategy to identify the optimal precursor represents a rather fundamental choice. In more applied examples one looks for precursors which minimize or maximize more sophisticated quantities, e.g., discriminant functions or loss matrices. These quantities are usually functions of the posterior PDF or the likelihood, but they take into account the additional demands of the specific problem, e.g., minimizing the loss due to a false prediction. The strategy studied in this contribution is thus fundamental in the sense that it enters into many of the more sophisticated quantities which were used for predictions and decision making.

B. Testing for predictive power

A common method to verify a hypothesis or to test the quality of a prediction is the receiver operating characteristic curve (ROC curve) [8,12,13]. The idea of the ROC curve consists simply of comparing the rate of correctly predicted events r_c with the rate of false alarms r_f by plotting r_c vs r_f. The rate of correct predictions r_c and the rate of false alarms r_f can be obtained by integrating the *aposterior* PDFs $\rho(\mathbf{x}_{(n-k+1,n)} | Y_{n+1}=1)$ and $\rho(\mathbf{x}_{(n-k+1,n)} | Y_{n+1}=0)$ on the precursory volume.

$$r_c(\delta, \mathbf{x}^{pre}) = \int \rho(\mathbf{x}_{(n-k+1,n)} | Y_{n+1} = 1) dV^{pre}(\delta, \mathbf{x}^{pre}), \quad (6)$$

$$r_f(\delta, \mathbf{x}^{pre}) = \int \rho(\mathbf{x}_{(n-k+1,n)} | Y_{n+1} = 0) dV^{pre}(\delta, \mathbf{x}^{pre}). \quad (7)$$

Note that these rates are defined with respect to the total number of events $Y_{n+1}=1$ and nonevents $Y_{n+1}=0$. Thus the relative frequency of events has no direct influence on the ROC curve, unlike on other measures of predictability, as, e.g., the Brier score or the ignorance [14].

Plotting r_c vs r_f for increasing values of δ one obtains a curve in the unit square of the r_f-r_c plane (see, e.g., Fig. 3). The curve approaches the origin for $\delta \rightarrow 0$ and the point $(1,1)$ in the limit $\delta \rightarrow \infty$, where δ accounts for the size of the precursory volume $V_{pre}(\delta)$. The shape of the curve characterizes the significance of the prediction. A curve above the diagonal reveals that the corresponding strategy of prediction is better than a random prediction which is characterized by the diagonal. Furthermore we are interested in curves which converge as fast as possible to 1, since this scenario tells us that we reach the highest possible rate of correct prediction without having a large rate of false alarms.

That is why we use the so-called *likelihood ratio* as a summary index, to quantify the ROC curve. For our inference problems the likelihood ratio is identical to the slope m of the ROC curve at the vicinity of the origin which implies $\delta \rightarrow 0$. This region of the ROC plot is in particular interesting, since it corresponds to a low rate of false alarms. The term likelihood ratio results from signal detection theory. In the context of signal detection theory, the term *a posterior PDF* refers to the PDF, which we call likelihood in the context of predictions and vice versa. This is due to the fact that the aim of signal detection is to identify a signal which was already observed in the past, whereas predictions are made about future events. Thus the "likelihood ratio" is in our notation a ratio of *a posteriori* PDFs.

$$m = \frac{\Delta r_c}{\Delta r_f} \sim \frac{\rho(\mathbf{x}^{pre}|Y_{n+1}=1)}{\rho(\mathbf{x}^{pre}|Y_{n+1}=0)} + \mathcal{O}(\delta). \tag{8}$$

However, we will use the common name likelihood ratio throughout the text. For other problems the name likelihood ratio is also used for the slope at every point of the ROC curve. Since we apply the likelihood ratio as a summary index for ROC curves, we specify that for our purposes the term likelihood ratio refers only to the slope of the ROC plot at the vicinity of the origin as in Eq. (8).

Note that one can show that the precursor, which maximizes the likelihood as explained in Sec. II A, also maximizes the m and is in this sense the optimal precursor.

C. Addressing the dependence on the event size

We are now interested in learning how the predictability depends on the event size η which is measured in units of the standard deviation of the time series under study. Thus the event variable Y_{n+1} becomes dependent on the event size,

$$Y_{n+1}(\eta) = \begin{cases} 1 & \begin{array}{l}\text{an event of size } \eta \text{ or larger} \\ \text{occurred at time } n+1 \end{array} \\ 0 & \begin{array}{l}\text{no event of size } \eta \text{ or larger} \\ \text{occurred at time } n+1. \end{array} \end{cases} \tag{9}$$

Via Bayes' theorem the likelihood ratio can be expressed in terms of the likelihood $L(Y_{n+1}(\eta)=1|\mathbf{x}_{pre})$ and the total probability to find events $P(Y_{n+1}(\eta)=1)$. Inserting the technical details of the calculation of the likelihood and the total probability (see the Appendix) we can see that the likelihood ratio depends sensitively on the joint PDF $j(\mathbf{x}_{(n-k+1,n)}, Y_{n+1}(\eta)=1)$ of pecursory variable and event.

$$m(Y_{n+1}(\eta), \mathbf{x}_{(n-k+1,n)}) = \frac{\left(1 - \int_{-\infty}^{\infty} d\mathbf{x}_{(n-k+1,n)} j(\mathbf{x}_{(n-k+1,n)}, Y_{n+1}(\eta)=1)\right)}{\int_{-\infty}^{\infty} d\mathbf{x}_{(n-k+1,n)} j(\mathbf{x}_{(n-k+1,n)}, Y_{n+1}(\eta)=1)} \frac{\dfrac{j(\mathbf{x}_{(n-k+1,n)}, Y_{n+1}(\eta)=1)}{\rho(\mathbf{x}_{(n-k+1,n)})}}{\left(1 - \dfrac{j(\mathbf{x}_{(n-k+1,n)}, Y_{n+1}(\eta)=1)}{\rho(\mathbf{x}_{(n-k+1,n)})}\right)},$$

with

$$j(\mathbf{x}_{(n-k+1,n)}, Y_{n+1}(\eta)=1) = \int_{\mathcal{M}} dx_{n+1} j(\mathbf{x}_{(n-k+1,n)}, x_{n+1}),$$

$$\mathcal{M} = \{x_{n+1}: Y_{n+1}=1\},$$

and

$$\rho(\mathbf{x}_{(n-k+1,n)}) = j(\mathbf{x}_{(n-k+1,n)}, Y_{n+1}(\eta)=1) \\ + j(\mathbf{x}_{(n-k+1,n)}, Y_{n+1}(\eta)=0). \tag{10}$$

Hence once the precursor is chosen, the dependence on the event size η enters into the likelihood ratio, via the joint PDF of event and precursor. Looking at the rather technical formula in Eq. (10), there are two aspects, which we find remarkable:

The slope of the ROC plot is fully characterized by the knowledge of the joint PDF of precursory variable and event. This implies that in the framework of statistical predictions all kinds of (long-range) correlations which might be present in the time series influence the quality of the predictions only through their influence on the joint PDF.

The definition of the event, e.g., as a threshold crossing or an increment does change this dependence only insofar as it enters into the choice of the precursor and it influences also the set on which the integrals in Eq. (10) are carried out. Both $Y_{n+1}(\eta)$ and the set $\mathcal{M}$ have to be defined according to the type of events one predicts. When predicting, e.g., increments $x_{n+1} - x_n \geq \eta$ via the precursory variable x_n, then $\mathcal{M}$

$=[a, b]$ with $a(Y_{n+1}(\eta)) = x_n + \eta$ for the lower border and $b(Y_{n+1}(\eta)) = \infty$ for the upper border. In order to predict threshold crossings at x_{n+1} via x_n one uses $a(Y_{n+1}(\eta)) = \eta$, $b(Y_{n+1}(\eta)) = \infty$.

Exploiting Eq. (10) we can hence determine the dependence of the likelihood ratio and the ROC curve on the events size η, via the dependence of the joint PDF of the process under study.

D. Constraint for increasing quality of predictions with increasing event size

In order to study the dependence of the likelihood ratio on the event size we are going to introduce a constraint which the likelihood and the total probability to find events have to fulfill in order to find a better predictability of larger (smaller) events.

In order to improve the readability of the paper, we will first introduce the following notations for the aposterior PDFs, the likelihood and the total probability to find events:

$$\rho_c(\mathbf{x}_{(n-k+1,n)}, \eta) = \rho(\mathbf{x}_{(n-k+1,n)}|Y_{n+1}(\eta)=1), \tag{11}$$

$$\rho_f(\mathbf{x}_{(n-k+1,n)}, \eta) = \rho(\mathbf{x}_{(n-k+1,n)}|Y_{n+1}(\eta)=0), \tag{12}$$

$$L(\eta, \mathbf{x}_{(n-k+1,n)}) = L(Y_{n+1}(\eta)=1|\mathbf{x}_{(n-k+1,n)}), \tag{13}$$

$$P(\eta) = P(Y_{n+1}(\eta)=1). \tag{14}$$

We can then ask for the change of the likelihood ratio with changing event size η.

$$\frac{\partial}{\partial \eta} m(Y_{n+1}(\eta), \mathbf{x}_{(n-k+1,n)}) \gtreqless 0. \tag{15}$$

The derivative of the likelihood ratio is positive (negative, zero), if the following sufficient condition $c(\eta)$ is fulfilled.

$$c(\eta, \mathbf{x}_{(n-k+1,n)}) = \frac{\partial}{\partial \eta} \ln L(\eta, \mathbf{x}_{(n-k+1,n)})$$
$$- \frac{(1 - L(\eta, \mathbf{x}_{(n-k+1,n)}))}{(1 - P(\eta))} \frac{\partial}{\partial \eta} \ln P(\eta) \gtreqless 0. \tag{16}$$

Hence one can tell for an arbitrary process, if extreme events are better predictable, by simply testing if the marginal PDF of the event and the likelihood of event and precursor fulfill Eq. (16).

III. PREDICTIONS OF INCREMENTS IN I.I.D. RANDOM NUMBERS

In this section we test the condition $c(\eta, \mathbf{x}_{(n-k+1,n)})$ as given in Eq. (16) for increments in Gaussian, power-law, and exponentially distributed i.i.d. random numbers. We thus concentrate on extreme events which consist of a sudden increase (or decrease) of the observed variable within a few time steps. Examples of this kind of extreme event are the increases in wind speed in [2,3], but also stock market crashes [15,16] which consist of sudden decreases.

We define our extreme event by an increment $x_{n+1} - x_n$ exceeding a given threshold η,

$$x_{n+1} - x_n \geq \eta, \tag{17}$$

where x_n and x_{n+1} denote the observed values at two consecutive time steps and the event size η is again measured in units of the standard deviation.

Since the first part of the increment x_n can be used as a precursory variable, the definition of the event as an increment introduces a correlation between the event and the precursory variable x_n. Hence the prediction of increments in random numbers provides a simple, but not unrealistic example which allows us to study the influence of the distribution of the underlying process on the event-size dependence of the quality of prediction.

In the examples which we study in this section the joint PDF of precursory variable and event is known and we can hence evaluate $c(\eta, x_n)$ analytically. A mathematical expression for a filter which selects the PDF of our extreme events out of the PDFs of the underlying stochastic process can be obtained through applying the Heaviside function $\Theta(x_{n+1} - x_n - \eta)$ to the joint PDF. This method is described in more detail in the Appendix.

Since in most cases the structure of the PDF is not known analytically, we are also interested in evaluating $c(\eta, x_n)$ numerically. In this case the approximations of the total probability and the likelihood are obtained by "binning and counting" and their numerical derivatives are evaluated via a Savitzky-Golay filter [17,18]. The numerical evaluation is done within 10^7 data points. In order to check the stability of

this procedure, we evaluate $c(\eta, x_n)$ also on 20 bootstrap samples which are generated from the original data set. These bootstrap samples consist of 10^7 pairs of event and precursory variable, which were drawn randomly from the original data set. Thus their PDFs are slightly different in their first and second moment and they contain different numbers of events. Evaluating $c(\eta, x_n)$ on the bootstrap samples thus shows how sensitive our numerical evaluation procedure is towards changes in the numbers of events. This is especially important for large and therefore rare events.

In order to check the results obtained by the evaluation of $c(\eta, x_n)$, we compute also the corresponding ROC curves analytically and numerically. Note that for both, the numerical evaluation of the condition and the ROC plots, we used only event sizes η for which we found at least 1000 events, so that the observed effects are not due to a lack of statistics of the large events.

A. Gaussian distributed random numbers

In the first example we assume the sequence of i.i.d. random numbers which form our time series to be normal distributed. As we know from [4], increments within Gaussian random numbers are better predictable the more extreme they are. In this section we will show that their PDFs fulfill also the condition in Eq. (16). Applying the filter mechanism developed in the Appendix we obtain the following expressions for the *a posteriori* PDFs:

$$\rho_c(x_n, \eta) = \frac{\exp\left(-\dfrac{x_n^2}{2\sigma^2}\right)}{2\sqrt{2\pi}\sigma P(\eta)} \operatorname{erfc}\left(\frac{x_n + \sigma\eta}{\sigma\sqrt{2}}\right), \tag{18}$$

and the likelihood

$$L(\eta, x_n) = \frac{1}{2} \operatorname{erfc}\left(\frac{x_n + \sigma\eta}{\sigma\sqrt{2}}\right). \tag{19}$$

We recall that the optimal precursor is given by the value of x_n which maximizes the likelihood. We refer to this special value of the variable x_n by x_{pre} and find for the likelihood according to Eq. (19) $x_{pre} = -\infty$. Thus, instead of a finite alarm volume δ here is the upper limit of the interval $[-\infty, \delta]$. The total probability to find increments of size η is given by

$$P(\eta) = \frac{1}{2} \operatorname{erfc}(\eta/2). \tag{20}$$

Hence the condition in Eq. (16) reads

$$c(\eta, x_n) = -\sqrt{\frac{2}{\pi}} \frac{\exp(-z^2)}{\operatorname{erfc}(z)}$$
$$+ \frac{1}{\sqrt{\pi}} \frac{\exp\left(-\dfrac{\eta^2}{4}\right)}{\operatorname{erfc}\left(\dfrac{\eta}{2}\right)} \frac{\left[1 - \dfrac{1}{2}\operatorname{erfc}(z)\right]}{\left[1 - \dfrac{1}{2}\operatorname{erfc}\left(\dfrac{\eta}{2}\right)\right]},$$

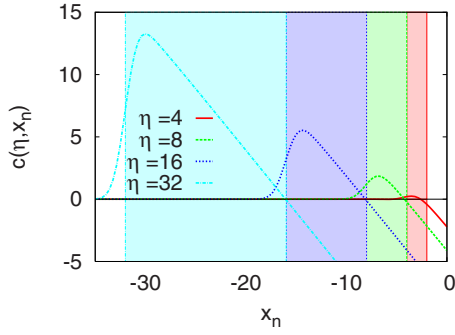with $\quad z = \dfrac{x_n + \sigma\eta}{\sqrt{2}\sigma}. \tag{21}$

FIG. 1. (Color online) The condition $c(\eta, x_n)$ for the Gaussian distribution as given by Eq. (21). The color shaded regions indicate the intervals $[-\sigma\eta, -\eta/2]$ for which we can, according to Eq. (23), expect $c(\eta, x_n)$ to be positive. If $x_n < -\sigma\eta$, $\eta > 2\sqrt{\pi}$ and terms of the order of $\exp[-(x_n+\sigma\eta)^2]$ are sufficiently small, the condition is also positive. If terms of the order of $\exp[-(x_n+\sigma\eta)^2]$ cannot be neglected one also might find small regions in $(-\infty, -\sigma\eta]$ for which $c(\eta, x_n) < 0$. However, the influence of these regions is neglectable, since our alarm interval is defined as $[-\infty, \delta]$ which implies an averaging over several possible values of the precursory variable.

Figure 1 illustrates this expression and Fig. 2 compares it to the numerical results. For the ideal precursor $x_n = x_{pre} = -\infty$ the condition $c(\eta, x_n)$ is—according to Eq. (21)—zero, since in this case, the slope of the ROC curve tends to infinity [4] and does not react to any variation in η. For any finite value of the precursory variable $x_n < 0$ we have to distinguish three regimes of $z = (x_n + \sigma\eta)/\sqrt{2}\sigma$, namely, $z \to \infty$ or $z \to -\infty$ and finally also the case $z = 0$.

In the first case we study the behavior of $c(\eta, x_n)$ for a fixed value of the precursory variable $-\sigma\eta < x_n$ and $\eta \to \infty$. This implies that $z \to \infty$ and we can use the asymptotic expansion for large arguments of the complementary error function
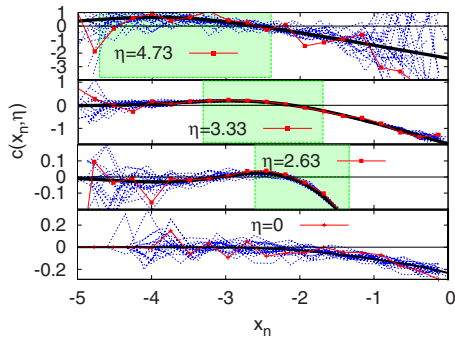

FIG. 2. (Color online) Comparison of the numerically evaluated condition $c(\eta, x_n)$ for the Gaussian distribution and the expression given by Eq. (21). The black curves denote the evaluation of the analytic result in Eq. (21), the curves plotted with lines and symbols represent the numerical results obtained from the original data set, and the dashed lines represent the results obtained from the corresponding bootstrap samples. The gray (green in the colored plot) regions indicate the regime $-\sigma\eta < x_n < -\sigma\eta/2$ for which $c(\eta, x_n)$ is positive in the limit $\eta \to \infty$. The numerical evaluation of $c(\eta, x_n)$ was done by sampling the likelihood and the total probability of events from 10^7 random numbers.
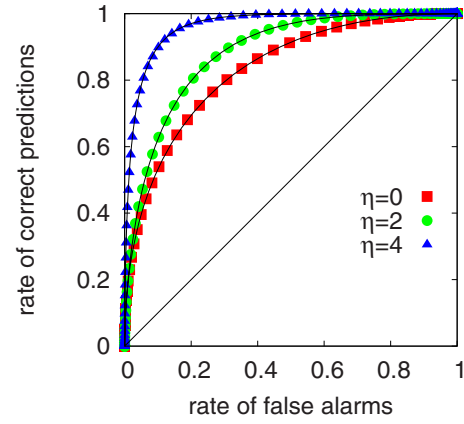

FIG. 3. (Color online) ROC plot for Gaussian-distributed i.i.d. random variables. The symbols represent ROC curves which where made via predicting increments in 10^7 normal i.i.d. random numbers. The predictions were made according to the prediction strategy described in Sec. II A. The lines represent the results of evaluating the integrals in Eqs. (6) and (7) for the Gaussian case. Note that the quality of the prediction increases with increasing event size.

$$\text{erfc}(z) \sim \frac{\exp(-z^2)}{\sqrt{\pi}z}\left(1 + \sum_{m=1}^{\infty}(-1)^m \frac{1 \times 3 \dots (2m-1)}{(2z^2)^m}\right)$$
$$\times \left(z \to \infty, |\arg z| < \frac{3\pi}{4}\right), \quad (22)$$

which can be found in [19] to obtain

$$c(\eta, x_n) \propto -\frac{x_n}{\sigma} + \frac{\eta}{2}, \quad -\sigma\eta < x_n < 0. \quad (23)$$

This expression is appropriate for $x_n > -\sigma\eta$ since the asymptotic expansion in Eq. (22) holds only if the argument of the complementary error function is positive. In this case $c(\eta, x_n)$ is larger than zero, if x_n is fixed and finite and $-\sigma\eta < x_n < -\sigma\eta/2$.

In the second case, we assume $\eta \gg 1$ to be fixed, $x_n < -\sigma\eta$, and $x_n \to -\infty$. Hence we can use the expansion in Eq. (22) only to obtain the asymptotic behavior of the dependence on η and not for the dependence on z. An asymptotic expression of $c(\eta, x_n)$ hence reads

$$c(\eta, x_n) \propto \frac{\eta}{2\left(1 - \frac{1}{2}\text{erfc}(\eta/2)\right)}\left(\frac{\text{erf}(z)}{\sqrt{pi}} + \frac{\eta}{2}\right)$$
$$- \mathcal{O}(\exp(-z^2)), \quad x_n < -\sigma\eta. \quad (24)$$

Since $\text{erf}(z)$ tends to minus unity as $z \to -\infty$ the expression in Eq. (24) is positive if $\eta > 2\sqrt{\pi}$ and if we can assume the squared exponential term to be sufficiently small. If the later assumption is not fullfilled one might observe some regions of intermediate values of $-\infty < x_n < -\sigma\eta$, for which $c(\eta, x)$ is negative.

However, the ROC curves in Fig. 3 suggest that the influe-

ence of these regions is sufficiently small, if the alarm volume is chosen to be $[-\infty, \delta]$. We can understand this effect, if we keep in mind that we use the interval $[-\infty, \delta]$ as an alarm volumen. Hence we can expect that the influence of the regions, where $c(\eta, x_n)$ is negative, is suppressed since we average over many different values of x_n and the condition is positive as $x_n \to -\infty$. (Positive is meant here in the sense that $c(\eta, x_n)$ approaches the value zero for $x_n = -\infty$ from small positive numbers.)

In the third case, for $x_n = -\sigma\eta$ and hence $z = 0$ we find that $c(\eta, x_n)$ is positive if $\eta > 2\sqrt{\frac{2}{\pi}}[1 - \frac{1}{2}\operatorname{erfc}(\eta/2)]$. In total we can expect larger increments in Gaussian random numbers to be easier to predict the larger they are. The ROC curves in Fig. 3 support these results.

B. Symmetrized exponential distributed random variables

The PDF of the symmetrized exponential reads

$$
\rho(x) = \frac{\lambda}{2}\exp(-\lambda|x_n|) = \begin{cases} \dfrac{\lambda}{2}\exp(-\lambda x_n), & x_n > 0 \\[2mm] \lambda/2, & x_n = 0 \\[2mm] \dfrac{\lambda}{2}\exp(\lambda x_n), & x_n < 0, \end{cases}
\tag{25}
$$

with $\mu = 0$, $\sigma = \sqrt{2}/\lambda$.

Applying the filtering mechanism according to the Appendix we find the joint PDFs of precursory variable and event

$$
j(x_n, (Y_{n+1}(\eta) = 1)) = \begin{cases} \dfrac{\lambda}{4}\exp(-\sqrt{2}\eta - 2\lambda x_n), & x_n > 0 \\[3mm] \dfrac{\lambda}{4}\exp(-\sqrt{2}\eta), & -\eta < x_n < 0 \\[3mm] \dfrac{\lambda}{2}\left[\exp(\lambda x_n) - \dfrac{1}{4}\exp(\sqrt{2}\eta + \lambda 2 x_n)\right], & x_n < -\eta < 0, \end{cases}
\tag{26}
$$

the aposterior probabilities,

$$
\rho_c(x_n, \eta, \lambda) = \begin{cases} \dfrac{\lambda}{(2 + \sqrt{2}\eta)}\exp(-2\lambda x_n), & x_n > 0 \\[3mm] \dfrac{\lambda}{(2 + \sqrt{2}\eta)}, & -\eta < x_n < 0 \\[3mm] \dfrac{\lambda}{(2 + \sqrt{2}\eta)}\left[2\exp(\sqrt{2}\eta + \lambda x_n) - \exp(2\sqrt{2}\eta + 2\lambda x_n)\right], & x_n < -\eta < 0, \end{cases}
\tag{27}
$$

$$
\rho_f(x_n, \eta, \lambda) = \begin{cases} \dfrac{\lambda}{2}\exp(-\lambda x_n)\dfrac{\left[1 - \frac{1}{2}\exp(-\lambda x_n - \sqrt{2}\eta)\right]}{\left[1 - \frac{1}{2}\left(1 + \frac{\eta}{2}\right)\exp(-\sqrt{2}\eta)\right]}, & x_n > 0 \\[6mm] \dfrac{\lambda}{2}\exp(\lambda x_n)\dfrac{\left[1 - \frac{1}{2}\exp(-\lambda x_n - \sqrt{2}\eta)\right]}{\left[1 - \frac{1}{2}\left(1 + \frac{\eta}{2}\right)\exp(-\sqrt{2}\eta)\right]}, & -\eta < x_n < 0 \\[6mm] \dfrac{\lambda}{4}\dfrac{\exp(2\lambda x_n + \sqrt{2}\eta)}{\left[1 - \frac{1}{2}\left(1 + \frac{\eta}{2}\right)\exp(-\sqrt{2}\eta)\right]}, & x_n < -\eta < 0, \end{cases}
\tag{28}
$$

the likelihood

$$L(\eta, x_n, \lambda) = \begin{cases} \frac{1}{2}\exp(-\sqrt{2}\eta - \lambda x_n), & x_n > 0 \\ \frac{1}{2}\exp(-\sqrt{2}\eta - \lambda x_n), & -\eta < x_n < 0 \\ 1 - \frac{1}{2}\exp(\sqrt{2}\eta + \lambda x_n), & x_n < -\eta < 0, \end{cases} \tag{29}$$

and the total probability to find events of size η,

$$P(\eta) = \begin{cases} \frac{1}{8}\exp(-\sqrt{2}\eta), & x_n > 0 \\ \frac{\sqrt{2}}{4}\eta\exp(-\sqrt{2}\eta), & -\eta < x_n < 0 \\ \frac{3}{8}\exp(-\sqrt{2}\eta), & x_n < -\eta < 0. \end{cases} \tag{30}$$

If we are not interested in the range of the precursory variable, the total probability to find events is given by

$$P(\eta) = \frac{1}{2}\exp(-\sqrt{2}\eta)\left(1 + \frac{\eta}{\sqrt{2}}\right). \tag{31}$$

Hence the condition $c(\eta, x_n, \lambda)$ reads

$$c(\eta, x_n, \lambda) = \begin{cases} -\sqrt{2}\left(1 - \frac{\left[1 - \frac{1}{2}\exp(-\sqrt{2}\eta - \lambda x_n)\right]}{\left[1 - \frac{1}{8}\exp(-\sqrt{2}\eta)\right]}\right), & x_n > 0 \\ -\sqrt{2} + \frac{(1 - \sqrt{2}\eta)}{\eta}\frac{\left[1 - \frac{1}{2}\exp(-\sqrt{2}\eta - \lambda x_n)\right]}{\left[1 - \frac{\sqrt{2}}{4}\eta\exp(-\sqrt{2}\eta)\right]}, & -\eta < x_n < 0 \\ -\frac{1}{\sqrt{2}}\exp(\lambda x_n + \sqrt{2}\eta)\left(\frac{1}{1 - \frac{1}{2}\exp(\sqrt{2}\eta + \lambda x_n)} + \frac{1}{1 - \frac{3}{8}\exp(-\sqrt{2}\eta)}\right), & x_n < -\eta. \end{cases} \tag{32}$$

Figure 4 compares the results of the numerical evaluation of the condition and the analytical expression given by Eq. (32). Since most precursors of large increments can be found among negative values, the numerical evaluation of $c(\eta, x_n\lambda)$ becomes worse for positive values of x_n, since in this limit the likelihood is not very well sampled from the data. This leads also to the wide spread of the bootstrap samples in this region.

Figure 4 shows that in the vicinity of the smallest value of the data set, the condition $c(\eta, x_n, \lambda)$ is zero. As we approach larger values of η, $c(\eta, x_n, \lambda)$ approaches zero in the whole range of data values. That is why we would expect to see no influence of the event size on the quality of predictions in the exponential case.

The ROC curves in Fig. 5 support these results. The numerical ROC curves were made via predicting increments in 10^7 normal i.i.d. random numbers according to the prediction

strategy described in Sec. II A. The precursor for the ROC curves is chosen as the maximum of the likelihood according to Eq. (29), i.e., $x_{pre} = -\infty$, so that the alarm interval is $[\infty, \delta]$. In summary there is no significant dependence on the event size for the prediction of increments in a sequence of symmetrical exponential distributed random numbers.

C. Pareto distributed random variables

We investigate the Pareto distribution as an example for power-law distributions. The PDF of the Pareto distribution is defined as [20]

$$\rho(x) = k x_{min}^k x^{-(k+1)} \tag{33}$$

for $x \in [x_{min}, \infty)$ with the exponent $k \geq 3$, the lower endpoint $x_{min} > 0$, and variance $\sigma = \frac{x_{min}}{k-1}\sqrt{\frac{k}{k-2}}$. Filtering for increments
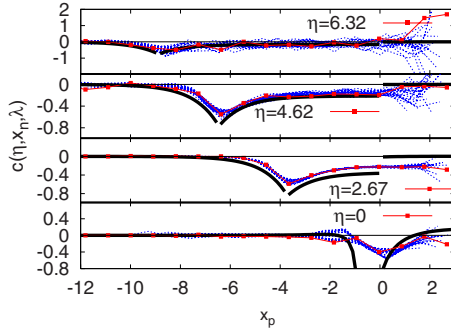
FIG. 4. (Color online) The numerically and analytically evaluated condition for the symmetrized exponential. The black line is the result of the analytical evaluation according to Eq. (32), the curves plotted with lines and symbols represent the numerical results obtained from the original data set, and the dashed lines represent the results obtained from the corresponding bootstrap samples. Note that for small values of x_n the condition $c(\eta, x_n, \lambda)$ is for all values of η close to zero.

of size η we find the following conditional PDFs of the increments:

$$\rho_c(x_n, \eta, x_{min}, k) = \frac{k x_{min}^{2k}}{x_n^{k+1}\left(x_n + \frac{x_{min}}{k+1}\sqrt{\frac{k}{k-2}}\eta\right)^k P(\eta, k)}, \quad (34)$$

$$\rho_f(x_n, \eta, x_{min}, k) = \frac{k x_{min}^k}{x_n^{k+1}} \frac{\left[1 - \left(\frac{x_{min}}{x_n + \frac{x_{min}}{k+1}\sqrt{\frac{k}{k-2}}\eta}\right)^k\right]}{1 - P(\eta, k)}, \quad (35)$$
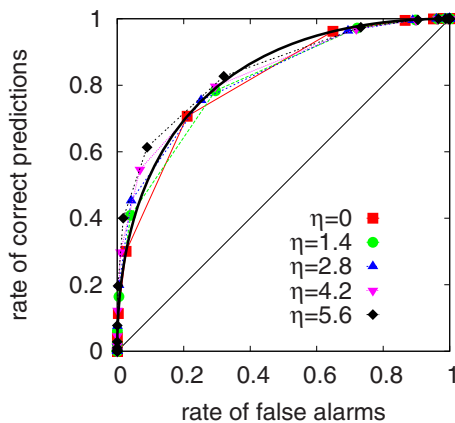

FIG. 5. (Color online) The ROC curves for symmetrically exponentially distributed i.i.d. random numbers show no significant dependence on the event size. The ROC curves where made via predicting increments in 10^7 normal i.i.d. random numbers and the predictions were made according to the prediction strategy described in Sec. II A. The black line indicates the analytically evaluated ROC curve for $\eta=0$.
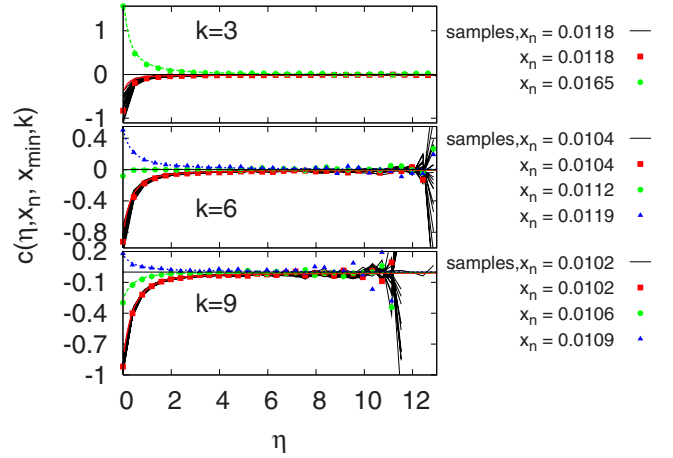

FIG. 6. (Color online) The condition $c(\eta, x_n, x_{min}, k)$ for the power-law distribution with lower endpoint $x_{min}=0.01$ are plotted for constant values of the precursory variable x_n. The symbols represent the results of the numerical evaluation of $c(\eta, x_n, x_{min}, k)$, the gray (colored) lines denote the analytic results, and the black lines denote the result for the corresponding bootstrap samples and the optimal precursor. For the "ideal" precursor $x_n = x_{min} = 0.01$ all values of $c(\eta, k, 0.01)$ are negative. Hence one should expect smaller events to be better predictable. However, this effect is sensitive of the choice of the precursor.

$$L(\eta, x_n, x_{min}, k) = \left(\frac{x_{min}}{x_n + \frac{x_{min}}{k+1}\sqrt{\frac{k}{k-2}}\eta}\right)^k. \quad (36)$$

Within the range (x_{min}, ∞) the likelihood has no well defined maximum. However, since the likelihood is a monotonously decreasing function, we use the lower endpoint x_{min} as a precursor. The total probability to find events of size η is given by

$$P(\eta, k) = \frac{1}{2}\,{}^2F^1\left(k, 2k, 2k+1, -\frac{\eta}{(k+1)}\sqrt{\frac{k}{k-2}}\right), \quad (37)$$

where ${}^2F^1(a, b, c, x)$ denotes the hypergeometric function $p_2F_q(a, b, c, x)$ with $p=2$, $q=1$.

Using

$$\frac{\partial P(\eta, k)}{\partial \eta} = \frac{k}{\eta}\left(\frac{1}{\left(1 + \frac{\eta}{k+1}\sqrt{\frac{k}{k-2}}\right)^k} - 2P(\eta, k)\right), \quad (38)$$

and inserting the expressions (36) and (34) for the components of $c(\eta, x_n, x_{min}, k)$ we can obtain an explicit analytic expression for the condition. In Fig. 6 we evaluate this expression using *Mathematica* and compare it with the results of an empirical evaluation on the data set of 10^7 i.i.d. random numbers.

Figure 6 displays that the value of $c(\eta, x_n, x_{min}, k)$ depends sensitively on the choice of the precursor. For the ideal precursor $x_{pre}=x_{min}$ all values of $c(\eta, k, x_{min})$ are negative. Hence one should in this case expect smaller events to be
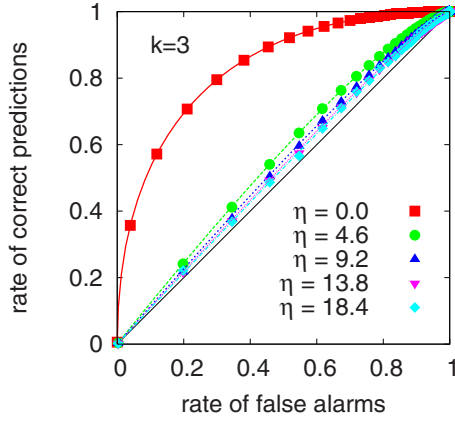
FIG. 7. (Color online) ROC plot for the power-law distribution with $k=3$ and $x_{min}=0.01$. The symbols show the numerical results and the lines indicate the analytically calculated ROC curves. The ROC curves were made via predicting increments in 10^7 Pareto distributed i.i.d. random numbers. The predictions were made according to the prediction strategy described in Sec. II A. Note that we tested only event sizes η, for which we found at least 1000 events, so that the effects we observe are not due to a lack of statistics of the large events. The ROC curves display that in Pareto distributed i.i.d. random numbers with the lower endpoint x_{min} $=0.01$ smaller events are better predictable and that large events are very hard to predict.

better predictable. The corresponding ROC curves in Figs. 7 and 8 verify this statement of $c(\eta, x_n, x_{min}, k)$.

In summary we find that larger events in Pareto distributed i.i.d. random numbers are harder to predict the larger they are. This is an admittedly unfortunate result, since extremely large events occur much more frequently in power-law distributed processes than in Gaussian distributed processes. Hence, their prediction would be highly desirable.
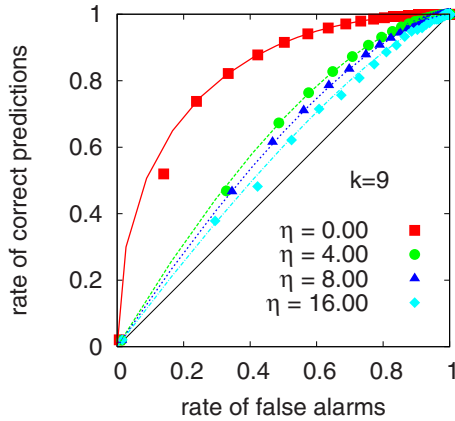

FIG. 8. (Color online) ROC plot for the power-law distribution with $k=9$ and $x_{min}=0.01$. The symbols show the numerical results and the lines indicate the analytically calculated ROC curves. The ROC curves were made via predicting increments in 10^7 Pareto distributed i.i.d. random numbers and the predictions were made according to the prediction strategy described in Sec. II A. The ROC curves display that in Pareto distributed i.i.d. random numbers smaller events are better predictable and large events are especially hard to predict.
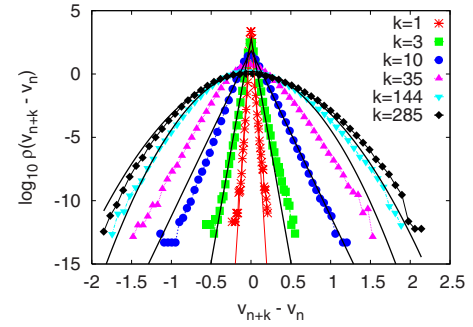

FIG. 9. (Color online) PDF of the increments $a_{n,k}=v_{n+k}-v_n$ with $k=1$, 3, 10, 35, 144, 285. The black lines correspond to Gaussian and exponential PDFs with appropriate values for the standard deviation or the coefficient λ.

IV. INCREMENTS IN FREE JET DATA

In this section, we apply the method of statistical inference to predict acceleration increments in free jet data. Therefore we use a data set of 1.25×10^7 samples of the local velocity measured in the turbulent region of a round free jet [21]. The data were sampled by a hot-wire measurement in the central region of an air into air free jet. One can then calculate the PDF of velocity increments $a_{n,k}=v_{n+k}-v_n$, where v_n and v_{n+k} are the velocities measured at time step n and $n+k$. The Taylor hypothesis allows one to relate the time resolution to a spatial resolution [21]. One observes that for large values of k the PDF of increments is essentially indistinguishable from a Gaussian, whereas for small k, the PDF develops approximately exponential wings [22–24]. Figure 9 illustrates this effect using the data set under study. Thus the incremental data sets $a_{n,k}$ provide us with the opportunity to test the results for statistical predictions within Gaussian and exponential distributed i.i.d. random numbers on a data set, which exhibits correlated structures.

We are now interested in predicting increments of the acceleration $a_{n+j,k}-a_{n,k} \geq \eta$ in the incremental data sets $a_{n,k}$ $=v_{n+k}-v_n$. In the following we concentrate on the incremental data set $a_{n,10}$, which has an asymptotically exponential PDF and the data set $a_{n,144}$, which has an asymptotically Gaussian PDF. Furthermore we focus on increments between relatively large time steps, i.e., $j=285$, so that the short-range persistence of the process does not prevent large events from occurring. As in the previous sections we are hence exploiting the statistical properties of the time series to make predictions, rather than the dynamical properties.

We can now use the evaluation algorithm which was tested on the previous examples to evaluate the condition for these data sets. The results are shown in Fig. 10. We find that at least for larger values of η the main features of $c(x_n, \eta)$ for the exponential and the Gaussian case as described in Secs. III A and III B are also present in the free jet data. For larger values of η, $c(a_{n,k}, \eta)$ is either larger than zero in the Gaussian case $(k=144)$ or equal to zero in the exponential case $(k=10)$ in the region of interesting precursory variables, i.e., small values of $a_{n,k}$.

However, the presence of the exponential and the Gaussian distributions is more prominent in the corresponding
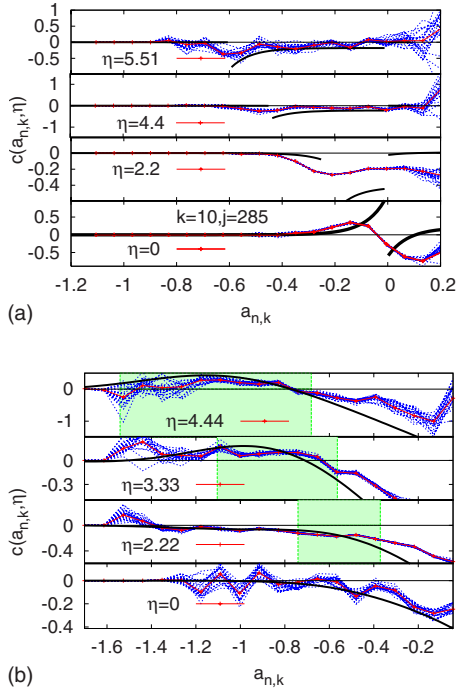
FIG. 10. (Color online) Transition from the exponential regime (a) to the Gaussian regime (b) characterized via the numerical evaluation of $c(x_n, \eta)$. The black line corresponds to the analytic results for the Gaussian and the exponential PDF, fitted to the PDFs of the increments, as it is shown in Fig. 9. For larger values of η the main features of $c(x_n, \eta)$ for the exponential and the Gaussian case as described in Secs. III A and III B are reproducible. For larger values of η we find that if $-\sigma\eta < a_{n,k} - \sigma\eta/2\ c(a_n, \eta)$ is either larger than zero in the asymptotically Gaussian case ($k=144$) or equal to zero in the asymptotically exponential case ($k=10$).

ROC curves. For the free jet data set, the predictions were made with an algorithm similar to the one described in Sec. II A. Instead of a specific precursory structure, which corresponds to the maximum of the likelihood, we use here a threshold of the likelihood as a precursor. In this setting we give an alarm for an extreme event, whenever the likelihood that an extreme event follows an observation is larger than a given threshold value.

In the exponential case ($k=10$) shown in Fig. 11(a) the ROC curves for different event size η almost coincide, although the range of η is larger $[\eta \in (0, 6.71)]$ than in the Gaussian case shown in Fig. 11(b). For $k=144$ the ROC curves are further apart, which corresponds to the results of Secs. III A and III B.

This example of the free jet data set shows that the specific dependence of the ROC curve on the event size can also in the case of correlated data sets be characterized by the PDF of the underlying process.

V. CONCLUSIONS

We study the size dependence of the quality of predictions for increments in a time series which consists of sequences of i.i.d. random numbers and in acceleration increments mea-
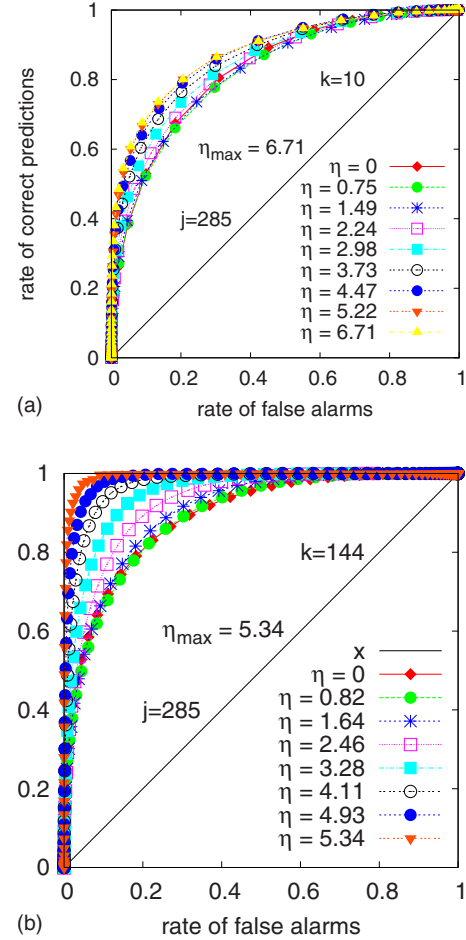

FIG. 11. (Color online) Transition from exponential ROC curves (a) to Gaussian ROC curves (b). In the exponential case ($k=10$), shown in (a) the ROC curves for different event size η are almost the same, although the range of η is larger $[\eta \in (0, 6.71)]$ than in the Gaussian case shown in (b). For $k=144$ the ROC curves are further apart, which corresponds to the results for Gaussian ROC curves (see Sec. III A).

sured in a free jet flow. Using the first part of the increment x_n as a precursory variable we predict large increments $x_{n+1} - x_n$ via statistical considerations. In order to measure the quality of the predictions we use ROC curves. Furthermore we introduce a quantitative criterion which can determine whether larger or smaller events are better predictable. This criterion is tested for a time series of Gaussian, exponential, and Pareto i.i.d. random variables and for the increments of the acceleration in the free jet flow. The results obtained from the criterion comply nicely with the corresponding ROC curves. Note that for both, the numerical evaluation of the condition and the ROC plots, we used only event sizes η for which we found at least 1000 events, so that the observed effects are not due to a lack of statistics of the large events.

In the sequence of Gaussian i.i.d. random numbers, we find that large increments are better predictable the larger they are. In the Pareto distributed time series we observe that in slowly decaying power laws larger events are harder to predict the larger they are. We find no significant dependence

on the event size for the sequence of exponentially i.i.d. random numbers.

While the condition can be easily evaluated analytically, it is not that easy to compute numerically from observed data, since the calculation implies evaluating the derivatives of numerically obtained distributions. Using Savitzky-Golay filters improved the results, but especially in the limit of larger events, where the distributions are difficult to sample, one cannot trust the results of the numerically evaluated criterion. However, it is still possible to apply the criterion by fitting a PDF to the distribution of the underlying process and then evaluate the criterion analytically.

Although the size dependence of the quality of predictions was observed in different contexts and for different measures of predictability, in this contribution only ROC curves were used. In order to exclude the possibility that the effect is specific to the ROC curve, future works should also include other measures of predictability.

Reviewing the results for the Gaussian case and the slowly decaying power law from a philosophic point of view one can conclude that nature allows us to predict large events from the most frequently occurring distribution easily. However, in Gaussian distributions very large events are rare and therefore less likely to cause damage. Whereas in the less frequently occurring distributions with heavy power-law tails, large events are especially hard to predict. Therefore one can assume that rare large impact events of processes with power-law distributions will remain unpredictable, although their prediction would be highly desirable.

APPENDIX: OBTAINING THE ANALYTIC EXPRESSION FOR THE LIKELIHOOD, THE JOINT AND THE APOSTERIOR PDFS FOR INCREMENTS IN STOCHASTICAL PROCESSES

An analytic expression for a filter which selects the PDF of our extreme increments $x_{n+1} - x_n \geq d$ out of the PDFs of the underlying stochastic process can be obtained through the Heaviside function $\Theta(x_{n+1} - x_n - d)$. (Note that d is not scaled by the standard deviation, i.e., $d = \sigma \eta$.) This filter is then applied to the joint PDF $j(x_0, x_1, \ldots, x_{n-k+1}, x_{n-k+2}, \ldots, x_n)$ of a stochastic process or to be more precise to the likelihood $L(x_{n+1} | x_0, x_1, \ldots, x_{n-k+1}, x_{n-k+2}, \ldots, x_n)$ that the $n+1$ step follows the previously obtained values. If we condition only on the last k values, we neglect the dependence on the past. The likelihood that an event $Y(d) = 1$ follows in the $n+1$th step can then be obtained by multiplication with $\Theta(x_{n+1} - x_n - d)$.

$$L(Y_{n+1}(d) = 1 | \mathbf{x}_{(n-k+1,n)}) = \Theta(x_{n+k} - x_n - d) L(x_{n+1} | \mathbf{x}_{(k,n)}), \tag{A1}$$

where $\mathbf{x}_{(n-k+1,n)} = (x_{n-k+1}, x_{n-k+2}, \ldots, x_n)$ as defined in Sec. II A. If the resulting expression is nonzero, the condition of the extreme event (17) is fulfilled and for x_{n+1} and x_n the following relation holds:

$$x_{n+1} = x_n + d + \gamma \quad (\gamma \in \mathbb{R}, \gamma \geq 0). \tag{A2}$$

Hence it is possible to express the likelihood in terms of x_n, which is a part of the precursory structure. We can use the integral representation of the Heaviside function with appropriate substitutions to obtain

$$L(Y_{n+1}(d) = 1 | \mathbf{x}_{(n-k+1,n)}) = \int_0^\infty L(x_n + d + \gamma | \mathbf{x}_{(n-k+1,n)}) d\gamma. \tag{A3}$$

Hence the joint PDF, the aposterior PDF, and the total probability to find increments are given by

$$j(\mathbf{x}_{(n-k+1,n)}, Y_{n+1}(d) = 1) = j(\mathbf{x}_{(n-k+1,n)}) \cdot L(Y_{n+1}(d) = 1 | \mathbf{x}_{(n-k+1,n)}), \tag{A4}$$

$$\rho(\mathbf{x}_{(n-k+1,n)} | Y_{n+1}(d) = 1) = \frac{j(\mathbf{x}_{(n-k+1,n)}, Y_{n+1}(d) = 1)}{P(Y(d) = 1)}, \tag{A5}$$

$$P(Y(d) = 1) = \int_{-\infty}^\infty dx_{n-k+1} \int_{-\infty}^\infty dx_{n-k+2} \ldots \int_{-\infty}^\infty dx_n j(\mathbf{x}_{(0,n-k)} \mathbf{x}_{(n-k+1,n)}, Y_{n+1}(d)) = 1. \tag{A6}$$

Whether we can access a given stochastical process analytically or not depends on the question of whether the integrals in Eq. (A6) can be solved or not.

If we are interested in the prediction of threshold crossings instead of increments, we can interpret η as the size of the threshold and set $x_n = 0$ in order to obtain the corresponding expressions for the likelihood, the joint PDF, the aposterior PDF, and the total probability.

[1] David D. Jackson, Proc. Natl. Acad. Sci. U.S.A. **93**, 3772 (1996).

[2] H. Kantz, D. Holstein, M. Ragwitz, and N. K. Vitanov, Physica A **342**, 315 (2004).

[3] Holger Kantz, Detlef Holstein, Mario Ragwitz, and Nikolay K. Vitanov, ShortTime Prediction of Wind Speeds from Local Measurements, in *Wind Energy—Proceedings of the EURO-MECH Colloquium*, edited by J. Peinke, P. Schaumann, and S. Barth (Springer, New York, 2006).

[4] S. Hallerberg, E. G. Altmann, D. Holstein, and H. Kantz, Phys. Rev. E **75**, 016706 (2007).

[5] M. Reza Rahimi Tabar, M. Sahimi, F. Ghasemi, K. Kaviani, M. Allamehzadeh, J. Peinke, M. Mokhtaru, M. Vesaghi, M. D. Niry, A. Bahraminasab, S. Tabatabai, S. Fayazbakhsh, and M. Akbari, e-print arXiv:physics/0510043.

[6] A. B. Shapoval and M. G. Shrirman, Int. J. Mod. Phys. C **17**, 1777 (2006).

[7] D. Lamper, S. D. Howison, and N. F. Johnson, Phys. Rev. Lett. **88**, 017902 (2001).

[8] J. P. Egan, *Signal Detection Theory and ROC Analysis* (Academic Press, New York, 1975).

[9] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis* (Prentice-Hall Inc., Englewood Clif., NJ, 1994).

[10] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods* (Springer, New York, 1998).

[11] J. M. Bernado and A. F. M. Smith, *Bayesian Theory* (Wiley-New York, 1994).

[12] D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics.* (Wiley, New York, 1966).

[13] M. S. Pepe, *The Statistical Evaluation of Medical Tests for Classification and Prediction*, (Oxford University Press, New York, 2003).

[14] J. Broecker and L. A. Smith, Weather Forecast. **22**, 382 (2007).

[15] A. Johansen and D. Sornette, Eur. Phys. J. B **1**, 141 (1998).

[16] N. Vandewalle *et al.* Eur. Phys. J. B **4**, 139 (1998).

[17] A. Savitzky and M. J. E. Golay, Anal. Chem. **36**, 1627 (1964).

[18] W. H. Press, *Numerical Recipes in C* (Cambridge University Press, Cambridge, England, 1992).

[19] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972).

[20] W. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1970), Vol. II.

[21] C. Renner, J. Peinke, and R. Friedrich, J. Fluid Mech. **433**, 383(2001).

[22] C. W. Van Atta and J. Park, in *Statistical Self-Similarity and Intertial Subrange Turbulence*, edited by M. Rosenblatt and C. W. Van Atta, Lect. Notes in Physics, Vol. 12 (Springer, Berlin, 1972). pp. 402–426.

[23] Y. Gagne, E. Hopfinger, and U. Frisch, in *New Trends in Nonlinear Dynamics and Pattern-Forming Phenomena*, edited by P. Coullet and P. Huerre, NATO Advanced Studies Institute (Plenum Press, New York, 1990), Vol. 237, pp. 315–319.

[24] U. Frisch, *Turbulence* (Cambridge University Press, Cambridge, England, 1995).

[25] In this contribution we use the name likelihood for the probability that an event follows a precursor **x** and the term *aposterior pdf* for the probability to find a precursor **x** before of an already observed extreme event. Note that the names might be also used vice versa, if one refers to the precursor as the previously observed information.